

# プロンプトに基づく俳句生成モデルの生成内容評価

佐藤 汰亮<sup>1</sup> 菊地 正弥<sup>1</sup> 大竹 孝樹<sup>1</sup> 今野 颯人<sup>1</sup> Ana Brassard<sup>2,1</sup>

大内 啓樹<sup>2</sup> 乾 健太郎<sup>1,2</sup>

<sup>1</sup> 東北大学 <sup>2</sup> 理化学研究所

{tasuku.sato.p6, seiya.kikuchi.t3}@dc.tohoku.ac.jp

{takaki, ryuto, inui}@ecei.tohoku.ac.jp

{ana.brassard, hiroki.ouchi}@riken.jp

## 1 はじめに

翻訳や対話応答生成などの文生成タスクは、sequence-to-sequence framework [1] (seq2seq) などのニューラルネットワークによって大きな進歩を遂げた。これらのタスクで生成される文（翻訳文、対話応答文など）は、入力する言語や内容に依存するものの、韻律や文字数といった言語の形式は比較的自由である。一方で、定型詩や歌詞は事前に決められた文字数や韻律などの制約に従って作成されるという特徴をもつ。そのため、これらのテキストの自動生成においては、モデルは事前に決められた制約を満たすような文を生成する必要がある [2, 3]。

本研究では、こういった制約のあるテキスト生成のタスクとして、俳句の自動生成に取り組む。俳句とは、上の句、中の句、下の句で構成された、定型詩の1つである。俳句は以下の3つの表層的な制約を満たしている必要がある。

- 5拍, 7拍, 5拍の合計 17 拍で構成されている
- 1つの季語が含まれている
- 切れ字が含まれている<sup>1)</sup>

制約のあるテキスト生成として俳句を選んだ理由は、その評価のしやすさにある。俳句は現在でも多くの人に親しまれている詩形式の1つであり、1つの句の長さが短く、切れ字や季語は判別が容易であるといった、生成結果の評価に際して都合のよい性質を持っている。我々は、表層的な制約を満たしつつ、句全体の意味が理解できる俳句の生成を目指す。

俳句の自動生成に関する既存の取り組みでは [4, 5, 6], 生成された俳句が表層的な制約を満たしているかについて評価が行われてきた。しかしなが

ら、生成された俳句が表層的な制約を満たしつつ、意味の理解できるものであるかについては十分に検討されていない。俳句のように表層的な制約を持つ詩の生成は、表層的な制約を満たすことに加え、意味を理解できる事が重要である [3]。例えば、「春の夜朝日輝く通りけり」という俳句は、上述の表層的な制約をすべて満たしているが、「夜に朝日が輝く」という意味の解釈が難しく、1つのまとまりのある句として理解・解釈ができないため、良い俳句とは言えない。

本研究では、俳句を自動生成するモデルを構築し、モデルによって生成された俳句が、表層的な制約を満たしつつ句全体の意味が理解できるものとなっているかを調査する。また、生成された俳句の内容に踏み込んだ考察をすることで、俳句生成への理解を深める。

## 2 関連研究

これまで、ルールベースやニューラルネットワークによって俳句を自動生成する研究がいくつか行われてきた。ルールベースによる俳句生成として土佐ら [4] が挙げられる。土佐らは、データベースを活用し、ルールによって俳句を生成するモデルを構築した。

ニューラルネットワークを用いた俳句生成には、太田ら [5] や米田ら [6] の研究がある。太田ら [5] は、キーワードを入力として俳句を生成する seq2seq モデルを提案した。太田らは、学習時に拍数や季節の素性を明示的に組み込む事で、俳句の表層的な制約を満たす俳句が生成されることを実現している。一方、米田ら [6] が行った俳句生成では、文生成モデルによって俳句を生成し、モチーフ画像に適合した俳句を選択することで画像と俳句を結びつけた。米田らは、「生成した俳句がどの程度訓練データと類

1) 切れ字を含むことを俳句の制約とするか否かは諸説ある。

表1 3つのプロンプトの具体例

入力	出力
[SEP] 冬天	に透く金の葉や樺の梢
鯉 [SEP]	群のまま寒鯉すこし動き けり
屋根 [SEP] 瓦 [SEP] 極月や葺	さいそせし屋根瓦

似しているか」という観点から、モデルの生成した俳句が表層的な制約を満たしているかの検証を行っている。

我々は、モデルによって生成された俳句について、表層的な制約を満たしているかという観点に加えて、句全体の意味が理解できるかという観点についても定量的に評価を行う。

### 3 入出力設定

本研究では、生成タスクで一般的に用いられている seq2seq を用いて俳句生成を行う。我々は、モデルによって生成される俳句をある程度制御するため、エンコーダーに俳句の元となるテキストを与え、与えたテキストを元にデコーダーによって俳句を自動生成する方法をとる。この方法には具体的に、(1) キーワードを入力として与え、キーワードを含んだ俳句を作成する方法と、(2) 上の句を入力として与え、それに続く中の句、下の句を作成する方法の2通りがあると考えられる。これらを踏まえ、俳句生成モデルの入出力として3つのプロンプトを定義した。

- 入力：俳句文頭の  $N$  文字， 出力：続きの俳句
- 入力：キーワード  $M$  個， 出力：キーワードを盛り込んだ俳句
- 入力：文頭の  $N$  文字とキーワード  $M$  個， 出力：キーワードを盛り込んだ続きの俳句

本稿で使用した学習データには、俳句の区切れ（上・中・下）の情報は含まれていないため、厳密に上の句を切り出すことができない。そこで、文頭の  $N$  文字を上上の句と仮定してモデルの入力とする。モデルの入力は、「キーワード」、「先頭  $N$  文字」の順に、これらを [SEP] トークンで連結した文字列とする。具体例を表1に示す。

## 4 実験設定

### 4.1 データセット

訓練及び評価に使用するデータとして、プロの詩人の作品や、コンテストにおける受賞作など、一定

表2 俳句生成モデルの構成

エンコーダのモデルアーキテクチャ	LSTM (bidirectional)
エンコーダの層数	4
エンコーダの隠れ層次元	512
デコーダのモデルアーキテクチャ	LSTM
デコーダの層数	4
デコーダの隠れ層次元	512
ミニバッチ最大トークン数	4000
最適化アルゴリズム	Adam[9] ( $\beta_1 = 0.9$ , $\beta_2 = 0.98$ , $\epsilon = 1 \times 10^{-8}$ )
学習更新回数	10,000
学習率	$5.0 \times 10^{-4}$
ドロップアウト率	0.3

の評価を得た俳句を集約した web データベースからデータをクロールし、俳句データセットを作成した。参照した web データベースは以下の4つである。(1) 一茶の俳句データベース（一茶俳句全集 V1.30）<sup>2)</sup>、(2) 松山市立子規念博物館正岡子規俳句<sup>3)</sup>、(3) 俳句例句データベース<sup>4)</sup>、(4) 現代俳句データベース<sup>5)</sup>。

3 節で説明した入出力のデータを作成するために、収集した俳句に対し、キーワードと文頭  $N$  文字を抽出する前処理を行った。俳句の文頭  $N$  文字は、人間が作成した 100 句の上の句の文字数をカウントして得られた離散分布に基づいてサンプリングした。サイトから俳句をスクレイピングした際に季語を取得できた俳句に関しては、季語をキーワードとして使用した。また、俳句を Mecab[7] で単語分割し、名詞と判定された単語もキーワードとして使用した。ただし、平仮名のみで構成された単語は除外した。 $M$  は  $1 \leq M \leq 3$  の整数値とした。トークナイズは文字分割を採用した。

以上の前処理を行い、最終的なデータセットの規模は、訓練データが約 386 万インスタンス、開発データが約 15 万インスタンスとなった。

### 4.2 モデル

俳句生成を行う seq2seq モデルのアーキテクチャは米田ら [6] と同様に LSTM を採用した。実装は fairseq[8] を使用した。モデルの更新回数は開発データを用いて決定した。モデルの詳細は表2に示す。

seq2seq モデルは、同じ語句を繰り返し生成して

2) <http://ohh.sisos.co.jp/cgi-bin/openhh/jsearch.cgi?group=hirarajp>

3) <http://sikihaku.lesp.co.jp/>

4) <http://taka.no.coocan.jp/a5/cgi-bin/HAIKUreikuDB/ZOU.htm>

5) <http://www.haiku-data.jp/kigo.html>

表3 生成俳句と重複度スコア

生成俳句	重複度スコア
林檎むく左右左右左右左右	0.43
雲の峰雲の中より雲の峰	0.52
松の葉の落ちたる松の落葉哉	0.65
青梅のひとつのことを思ひをり	0.75
晩年や涙のごとく秋の風	0.89

しまうことがある [10]. このような繰り返し生成をフィルタリングするため, 生成された俳句における単語分割での異なり語彙率と, 文字分割での異なり語彙率の平均値を重複度スコア  $S_{dupl} \in [0, 1]$  として定義した. ここで, 異なり語彙率とは, 異なり語数を総語数で割ったものである. 重複度スコアは, 同じ語句の繰り返しが多いほど低下する. 生成された俳句と重複度スコアの具体例を表3に示す. 開発データを用いて生成された俳句を元に閾値を0.7に設定し, 重複度スコアが閾値を下回った俳句は出力候補から除外した.

ビーム幅10のビームサーチによって生成された俳句のうち, 重複度スコアが閾値を上回った俳句からランダムに1つ選び, モデルの出力とした. ここで重複度スコアが最も高い俳句ではなく, ランダムに俳句を選択している理由は, 重複度スコアが同じ語句の繰り返し生成を抑制するために導入したスコアであり, 俳句の内容の質を反映したスコアではないためである. 条件を満たす俳句が存在しない場合, 重複度スコアが最も高い俳句を出力した.

3節で述べたように, モデルの入力としてキーワードが与えられる場合がある. このような入力に対しては, 入力されたキーワードを盛り込んだ俳句を出力することが望まれるため, 指定したキーワードが出力結果に含まれているかを検証した. 開発データのうち, キーワードを入力として与えているデータを用いて俳句を生成したところ, 97%の俳句で指定した全てのキーワードが含まれていることを確認した.

### 4.3 評価方法

1句につき3人のアノテーターが評価を行う. 各質問には回答の選択肢が2つ用意されており, 3人のうち2人が選択した選択肢をその俳句の評価として採用する. また, 各質問において, 3人のアノテーター間の一致率を fleiss kappa を用いて計算する.

表4 表層的な制約の評価

評価観点	割合	fleiss kappa
5/7/5 を満たしているか	82.0%	0.639
季語を含んでいるか	96%	0.277
切れ字を含んでいるか	32%	0.778

人手評価は俳句の表層的な制約と俳句の内容, 2つの観点に対して行う. 表層的な制約に関しては, 以下の3点を評価する.

- 5, 7, 5 の17音の制約を満たしているか
- 季語を含んでいるか
- 切れ字を含んでいるか

内容に関しては, 以下の2点を評価する.

- モデルの生成した俳句が意味を理解・解釈できる俳句となっているか
- (アノテーターが) 人が作成した俳句とモデルの生成結果, どちらの俳句を好むか

内容に関する2点目の評価項目については, (A) 人が作成した俳句, (B) (A) の俳句から抽出したキーワードと先頭  $N$  文字を入力としてモデルが生成した俳句, の2つを提示することで評価を行った. 評価時にはどちらがモデルの生成かは分からない状態にした. また, 俳句の良し悪しを客観的に判断することは難しいため, 「どちらの俳句が良い俳句か」ではなく「どちらの俳句が好きか」という質問を採用した.

人手評価用データは, 開発データを元に生成した俳句からランダムに50句サンプリングすることで作成した.

## 5 評価

### 5.1 表層的な制約の評価

表4に結果を示す. 拍数の制約は82%の割合で満たしており, アノテーター間での一致率は0.639だった. この評価では「5/7/5 を満たしているか」という問い方をしたため, 字余り, 字足らずの俳句においてアノテーター間の不一致が見られた. この様な字足らず, 字余りを許容すると, 拍数の制約を満たしている俳句の割合はより大きくなると考えられる.

季語は96%の割合で含まれており, 季語を含めなければいけないという制約を学習できていることがわかった.

表5 内容の評価

評価観点	割合	fleiss kappa
意味を理解・解釈できる	74.0%	0.179
モデルが生成した俳句が好き	42.0%	0.144

表6 俳句生成モデルの出力例

入力		
キーワード	先頭 N 文字	出力
墨, 夏, 香	-	墨の香のしづかに夏を 惜みけり
-	ふらここや	こころの中の水の音
女	待針の	女ばかりや秋の暮
-	団栗を	拾ふ峠のうすみどり

切れ字を含む確率は32%であった。切れ字を含むことを俳句の制約と考えるか否かは諸説あり、切れ字を含まない俳句も数多く存在する。実際、本研究で使用した訓練データ中にも切れ字を含まない俳句が見受けられる。ここから、切れ字のように、訓練データ中の俳句が必ず満たしているわけではない制約は、暗黙的に学習することが難しいと考えられる。

## 5.2 内容の評価

表5に結果を示す。fleiss kappaの値から、2つの質問はアノテーター間でも判断が分かれる観点であったと考えられる。人の作成した俳句よりもモデルが生成した俳句が好きだと回答した人は42%であった。また、意味を理解・解釈できる俳句は74%であった。理解、解釈できないと判断された生成俳句を確認すると、同じ単語が繰り返し出現しているものが散見された。ここから、人が理解・解釈できる俳句を生成するためには、同じ単語の繰り返し生成を抑制するための工夫を行うべきであると考えられる。

## 6 生成例・考察

表6に、俳句生成モデルが実際に生成した俳句を示す。俳句に求められる表層的な制約を満たしつつ、内容を理解・解釈できる俳句が生成されていることがわかる。

表7に、俳句の表層的な制約は満たしているものの、その内容の理解が難しい例を示す。

1行目は、同じ単語を繰り返し生成してしまっている例である。このような生成はseq2seqによくみられる現象である。

2行目は、常識を欠いた表現を生成している例である。常識を考えると、「膝に手をさしいれる」と

表7 内容の理解・解釈が難しい生成例

入力		
キーワード	先頭 N 文字	出力
もち	-	の花もち ちちちちちち ちちちち
膝	-	膝に手を さし入れてゐる 春の風邪
スマホ	-	クリスマス スマホとなりて ゐたりけり

いう表現には違和感がある。しかし、「手をさしいれる」が「風邪によって関節のリンパ節が痛くなる」ことを表現していると考え、風邪が自分の体を蝕んでいる様子を巧みに表現した俳句と見なすこともできる。俳句は書き手の表現力が要求されると同時に、読み手の想像力によっても作品の価値が高まる。よって、俳句生成時に常識を考慮していないことが俳句の質の悪化につながっているとは一概に言い切れない。

3行目は単語の意味を理解できずに生成している例である。これは、「スマホ」のように、学習データに含まれていない単語は、正しく生成結果に盛り込めないことや、俳句を文字分割して学習させているため、単語レベルの情報をモデルに組み込むことができていないことが原因と考えられる。

## 7 おわりに

本論文では、制約のあるテキスト生成として俳句生成に取り組んだ。プロンプトに基づく俳句生成モデルを構築し、その生成結果を表層的な制約、内容の2観点から人手評価した。とくに、「表層的な制約を満たす一方で、まとまりのある句として内容が理解できる俳句か？」という点に注目し、大部分の生成結果が表層的な制約を満たしている一方で、74%がその内容も理解・解釈できるものとなっていることを明らかにした。

今後の研究としては、繰り返し生成への対処、創作支援システムへの応用などが考えられる。特に、創作支援システムとして応用するために、入力形式の充実化に取り組みたい。また、俳句生成モデルが実際に創作支援ツールとして有効なのかを検証するための人手評価を行いたい。

**謝辞** 本研究はJSPS科研費JP19H04425の助成を受けたものである。

## 参考文献

- [1]Ilya Sutskever, Oriol Vinyals, and Quoc V. Le. Sequence to sequence learning with neural networks. In Zoubin Ghahramani, Max Welling, Corinna Cortes, Neil D. Lawrence, and Kilian Q. Weinberger, editors, *NeurIPS*, pp. 3104–3112, 2014.
- [2]Xingxing Zhang and Mirella Lapata. Chinese poetry generation with recurrent neural networks. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 670–680, Doha, Qatar, October 2014. Association for Computational Linguistics.
- [3]Piji Li, Haisong Zhang, Xiaojiang Liu, and Shuming Shi. Rigid formats controlled text generation. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pp. 742–751, Online, July 2020. Association for Computational Linguistics.
- [4]Naoko Tosa, Hideto Obara, and Michihiko Minoh. Hitch haiku: An interactive supporting system for composing haiku poem. In Scott M. Stevens and Shirley J. Saldamarco, editors, *Entertainment Computing - ICEC 2008*, pp. 209–216, Berlin, Heidelberg, 2009. Springer Berlin Heidelberg.
- [5]瑠子太田, 裕之進藤, 裕治松本. 深層学習を用いた俳句の自動生成. Technical Report 1, 奈良先端科学技術大学院大学, 奈良先端科学技術大学院大学, 奈良先端科学技術大学院大学, may 2018.
- [6]Koki YONEDA, Soichiro YOKOYAMA, Tomohisa YAMASHITA, and Hidenori KAWAMURA. Development of automatic haiku generator using lstm. *Proceedings of the Annual Conference of JSAI*, Vol. 2018, pp. 1B2OS11b01–1B2OS11b01, 2018.
- [7]Taku Kudo, Kaoru Yamamoto, and Yuji Matsumoto. Applying conditional random fields to japanese morphological analysis. In *EMNLP*, pp. 230–237, 2004.
- [8]Myle Ott, Sergey Edunov, Alexei Baevski, Angela Fan, Sam Gross, Nathan Ng, David Grangier, and Michael Auli. fairseq: A fast, extensible toolkit for sequence modeling. In *NAACL*, pp. 48–53, 2019.
- [9]Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In Yoshua Bengio and Yann LeCun, editors, *ICLR*, 2015.
- [10]Ari Holtzman, Jan Buys, Maxwell Forbes, and Yejin Choi. The curious case of neural text degeneration. *CoRR*, Vol. abs/1904.09751, , 2019.